(51) International Patent Classification [6] :		(11) International Publication Number:	**WO 99/11801**
C12N 15/63, 1/19, C12Q 1/37, C12N 9/50	**A2**	(43) International Publication Date:	11 March 1999 (11.03.99)

(21) International Application Number: PCT/GB98/02596

(22) International Filing Date: 28 August 1998 (28.08.98)

(30) Priority Data:
9718591.2   3 September 1997 (03.09.97)   GB

(71) Applicant (for all designated States except US): ZENECA LIMITED [GB/GB]; 15 Stanhope Gate, London W1Y 6LN (GB).

(72) Inventors; and
(75) Inventors/Applicants (for US only): BROAD, Peter, Michael [GB/GB]; Alderley Park, Macclesfield, Cheshire SK10 4TG (GB). CHARLES, Andrew, David [GB/GB]; Alderley Park, Macclesfield, Cheshire SK10 4TG (GB). HOLLIS, Melvyn [GB/GB]; Alderley Park, Macclesfield, Cheshire SK10 4TG (GB). MacCALLUM, Linda, Jean [GB/GB]; Alderley Park, Macclesfield, Cheshire SK10 4TG (GB). SCANLON, David, John [GB/GB]; Alderley Park, Macclesfield, Cheshire SK10 4TG (GB).

(74) Agent: BROWN, Andrew, Stephen; Zeneca Pharmaceuticals, Intellectual Property Dept., Mereside, Alderley Park, Macclesfield, Cheshire SK10 4TG (GB).

(81) Designated States: CA, IL, JP, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: PROTEASE BASED GENE SWITCHING SYSTEM

(57) Abstract

The present invention relates to materials and methods for protease–based gene switching systems, wherein a transcription factor is bound to a membrane via protease cleavage site. It also relates to the use of such materials and methods in the identification of substrates and inhibitors of proteases and in the design of altered specificity proteases.

PROTEASE BASED GENE SWITCHING SYSTEM

The present invention relates to materials and methods for protease-based gene switching systems. It also relates to the use of such materials and methods in the

5 identification of substrates and inhibitors of proteases and in the design of altered specificity proteases.

Proteases are involved in both intracellular and extracellular processes. Proteases are also essential in the propagation of pathogenic organisms. For these reasons, proteases are important targets for therapeutic agents. For example, inhibitors of angiotensin-converting

10 enzyme, such as captopril, have been used since the early 1980s in the treatment of hypertension (Materson and Preston, 1994). Amongst infectious agents, the identification of human immunodeficiency virus (HIV) protease inhibitors has led to new therapies for HIV infection (Richman, 1996).

Proteases may be useful agents in themselves. Delivery of a gene encoding a protease

15 to a target cell could result in the cleavage and alteration in activity of selected proteins. Proteases may be used in industrial or pharmaceutical processes to generate mature proteins or degrade undesirable proteins.

In general, two approaches have be taken to the design of assays which may be used to examine the properties of proteases. In the first approach an *in vitro* assay is used. This

20 requires a sample of the protease, isolated from either a natural source, or expressed from the gene encoding the protease in a heterologous system. In the second approach, a recombinant cell is configured so that some easily measurable property of a cell is made dependent upon the activity of a protease expressed within a cell. Such genetic systems for proteases are generally designed by incorporating a cleavage site for the protease into a target protein so

25 that when the protease cleaves the target protein, the function of the target protein is lost. For example, McCall *et al.* (1994) introduced a cleavage site for human rhinovirus protease 3C into a protein conferring tetracycline resistance. The modified protein is active and allows *Escherischia coli* (*E. coli*) cells to grow in the presence of tetracycline, unless the tetracycline resistance protein is cleaved by the protease. The cells which express active protease will not

30 grow in the presence of tetracycline. Baum *et al.* (1990) introduced a cleavage site for the HIV protease into the coding region of β-galactosidase in such a way that the enzyme retained

activity unless cleaved by the viral protease. Thus β-galactosidase activity is low in a cell
containing both the protease and the modified enzyme, but high in a cell which contains only
the modified enzyme. Protease cleavage sites have also been introduced into transcription
factors so that cleavage of the transcription factor results in the loss of transcriptional

5 activation capacity through separation of the activation domain and DNA binding domain
(Smith and Kohorn, 1991; DasMahaptra *et al.*, 1992). An analogous approach has been taken
to λ repressor, a prokaryotic gene regulator (Sices and Kristie, 1998). An HIV-1 protease site
was inserted between the DNA binding domain and the dimerisation domain. This modified
λ repressor is functional, but is non-functional when cleaved by the HIV-1 protease.

10 An alternative approach to the design of genetic systems for proteases is suggested in
Hirowatari (1995) which is to use protease cleavage to activate, rather than abolish, a property
of the substrate protein by releasing a transcription factor from an inactive membrane-bound
precursor. In some natural systems transcription factors are synthesised in an inactive form,
where inactivation is due to the association of the transcription factor with a membrane

15 (reviewed by Pahl and Baeuerle, 1996). The transcription factor SREBP-1 activates genes
involved in sterol biosynthesis. This protein is naturally synthesised with an amino terminal
extension which anchors it to the membrane of the endoplasmic reticulum (Wang *et al.*,
1994). The release of this transcription factor from the membrane is regulated by sterols and
requires two proteolytic cleavages within the membrane domain (Sakai *et al.*, 1996). In

20 Hirowatari *et al.* (1995) a chimaeric protein was constructed from two parts, in one part a
membrane anchor and a cleavage site for the HCV NS3 protease and in the second part the
transcriptional activator (Tax-1) from human T cell leukaemia virus type-1 (HTLV-1). A
reporter gene responsive to Tax-1 was therefore able to be activated when the HCV protease
and the chimaeric substrate protein where both present in the cell. In both of the above

25 examples the membrane anchor is derived from the same protein as the protease cleavage site.
We have found that it is possible to inactivate a transcription factor by attaching it to a
membrane anchoring domain via a protease cleavage site using completely unrelated
components. This considerably extends the utility of such systems in that it allows proteases
which do not normally cleave proteins with a membrane-bound precursor to be studied. Such

30 systems may be generally described as depicted in Figure 1. We have also found that such
systems have wider utility than previously disclosed, such as use as a gene switch, cloning

proteases, changing protease specificity and in gene therapy. The basic function of the system and the ease in which it may be used can be compared with the widely used yeast two-hybrid assay system. In the two-hybrid system two fusion proteins are expressed, preferably in a yeast cell, so that interaction of the two proteins results in activation of a target gene (Fields

5 and Song, 1989). In the system described herein, two proteins are also expressed, a protease and a substrate fusion protein. Cleavage of the substrate fusion protein by the protease results in target gene activation. In both cases the event of interest leads to target gene activation, and in both cases the systems may be configured in yeast cells, which are particularly amenable to genetic manipulation and for use in screening (Sherman, 1991).

10

Therefore in a first aspect of the invention we provide a heterologous cell which comprises:

(i)   a transcription factor precursor which comprises a transcription factor linked to a membrane anchoring domain via a protease cleavage site, in which the membrane

15   anchoring domain and protease cleavage site are not derived from the same protein;

(ii)   a protease which recognises the protease cleavage site in the transcription factor precursor; and

(iii)   a target gene under the control of the transcription factor, wherein if cleavage of the protease cleavage site by the protease is allowed to occur subsequent release of the

20   transcription factor enhances expression of the target gene.

In a cell configured in this way the target gene is not expressed, or is expressed at low levels, when cleavage of the transcription factor precursor is prevented, but when cleavage is allowed to proceed the transcription factor is released and expression of the target gene, or

25 genes, is measurably increased. As will be seen below there are many points within the configured cell at which cleavage of the protease cleavage site may be blocked. For instance the protease cleavage site may be altered, the specificity of the protease may be altered, or a molecule may interfere between the protease and the protease cleavage site. The above system is constructed in such a manner that the effects of any such changes may be directly measured

30 by monitoring expression of the target gene.

- 4 -

A particularly useful membrane localisation domain is the amino terminal region of the enzyme hydroxymethylglutaryl coenzyme A reductase (HMG-CoA reductase). HMG-CoA reductase from *Saccharomyces cerevisiae* (*S.cerevisiae*) has an amino terminal region with seven transmembrane helices (Basson *et al.*, 1988). Other proteins which contain

5   different numbers of transmembrane helices, such as members of the HMG-CoA reductase family from other organisms (Hampton *et al.*, 1996) could also be used, provided that the fused transcription factor is exposed to the cytosolic or nuclear compartment so that the released transcription factor does not need to be translocated across a cellular membrane to reach the target gene. As an alternative to peptide membrane anchors, lipid membrane

10 anchors may be used. For example, peptide sequences which are substrates for myristoylation (Pellman *et al.*, 1985) or farnesylation and farnesylation-dependent modification (Hancock *et al.* , 1991) may be used in place of transmembrane peptide domains.

The transcription factor can be a natural transcription factor, a chimaeric factor containing functional domains from different proteins, for example a DNA binding domain

15 linked to an activation domain, or it may contain synthetic domains. DNA binding domains which may be used include those of the *S. cerevisiae* factor Gal4 (Keegan *et al.*, 1986), the *E. coli* protein LexA (Brent and Ptashne, 1981) and a variety of other proteins such as those listed by Harrison (1991). Transcriptional activation domains include naturally-occurring domains, such as the activation domain of herpes simplex virus VP16 protein (Triezenberg *et*

20 *al.*, 1988) and the activation domains of Gal4 (Ma and Ptashne, 1987a) as well as acidic domains generated from semi-random sequence libraries (Ma and Ptashne, 1987b). Other activation domains have been listed by Triezenberg (1995). The transcription factor does not need to bind directly to DNA but may instead bind to proteins which do bind to DNA. The transcription factors, or proteins to which the transcription factors bind, will bind to

25 promoter/enhancer sequences upstream from the target gene. Such sequences are well known in the literature and are described in the above references.

The cell types within which such a system may be configured include: prokaryotic cells (*E. coli*); eukaryotic cells such as those of the model organisms budding yeast (*S. cerevisiae*), fission yeast *(Schizosaccharomyces pombe)*, the fruit fly (*Drosophila*

30 *melanogaster*), the nematode worm (*Caenorhabditis elegans*) and the plant *Arabidposis*

thaliana; maize (*Zea mays*); and mammalian cells such as primary human cells, established human cell lines, primary mouse cells and established mouse cell lines.

The target gene used to measure the output of the system may be one which produces an easily measurable gene product (a target or reporter gene) such as *E. coli* β-galactosidase

5   (Casadaban *et al.* 1983), firefly luciferase (de Wet *et al.*, 1987), *E. coli* chloramphenicol acetyl transferase (Gorman *et al.*, 1982), *E.coli* β-lactamase (Zlokarnik *et al.*, 1998) or green fluorescent protein (Chalfie *et al.*, 1994). In an alternative aspect of the invention the target gene may be a toxic gene, and activation of this gene through the action of the protease will result in cell death.

10 Rather than the target gene being used as a measure of the success in whether the protease has cleaved the substrate fusion protein the target gene may be a useful gene which requires regulation. In this aspect of the invention we may describe the target gene as being under the control of a protease-dependent gene switch. In general a gene switch is a system in which a gene of interest is turned off or is largely inactive, or alternatively is turned on, in one

15 state, but may be switched on, or off, by some alteration in the cell. One type of gene switch employs small molecules to effect the switch, for example, tetracycline-regulatable gene switches have been described in mammalian cells (see Shockett and Schatz, 1996). In these systems, the addition of tetracycline can either activate or repress a gene, depending on how the system is configured. We describe here a different type of gene switch in which the

20 switching event is effected by a protease. The advantage of this system is the high degree of control which may be obtained; in the absence of protease or where the activity of the protease is blocked, the target gene can be effectively completely inactive. In the presence of the protease, or absence of an inhibitor, it can be activated to a high level. Control of the switch may be exerted in the expression or the activity of the protease, for example in the use of a

25 modulator (inhibitor or activator) of the protease, or a modulator of the level of expression of the protease. In the latter case it would be possible, for example, to place the gene encoding the protease under the control of a tetracycline-regulatable promoter and achieve protease regulation through tetracycline. This may be superior to regulating the target gene with tetracycline itself, since the insertion of the protease system between the tetracycline-

30 regulatable promoter and the target gene introduces an amplification step.

Therefore in a further aspect of the invention we provide a gene switch mechanism comprising:

(i)   a transcription factor linked to a membrane anchoring domain via a protease cleavage site;

5 (ii)   a protease which recognises the protease cleavage site in the transcription factor precursor; and

(iii)   a gene placed under the control of the transcription factor, whereby enhanced expression of the gene occurs after cleavage of the protease cleavage site by the protease and thereby expression of the gene may be modulated by directly or indirectly

10   affecting the activity or expression of the protease.

A further aspect of the invention relates to the possibility of identifying peptide substrates for a protease using a cell configured in accordance with the invention. Identifying the peptide substrates for proteases is useful in that (i) they provide information which may

15 lead to the discovery of the substrate protein, thus elucidating the biological function of the protease (ii) they provide substrates which can be used in *in vitro* assays to screen for inhibitors of the proteases (iii) derivatives of the substrates may in themselves be inhibitors of the proteases.

With the complete sequencing of several microbial genomes (see Doolittle, 1997), the

20 complete sequencing of a model eukaryote (the budding yeast *S. cerevisiae*: Goffeau *et al.*, 1996) and the ongoing sequencing of expressed sequences and genomic sequences from the human genome, a number of open reading frames have been identified which, on the basis of homology searches, are likely to be proteases. The biological functions of these proteases cannot be understood until the substrates upon which they act are identified. One example of

25 such a family of protease is the metalloproteinase-disintegrin family. The family currently contains more than 20 members (Blobel, 1997). This family includes the enzyme TNFα-converting enzyme, considered to be a target for therapeutic agents in inflammation (Black *et al.*, 1997; Moss *et al.*, 1997). However, the substrate proteins of most members of this family remain unknown (Blobel, 1997).

30   One route to the identification of peptide substrates is to select from a library of potential substrates the sequences which are substrates for the protease. This allows the

definition of a consensus sequence for the cleavage site for the protease. Using bioinformatic search tools, databases of protein sequences can then be searched for sequences which correspond to this consensus. The existence of the consensus sequence within a protein will indicate that the protein is a potential substrate for the protease. Further evidence can then be

5 used to assess the likelihood that the protein identified is a substrate for the protease; for example whether they are expressed temporally and spatially in a way that would allow the protease to act upon the substrate. Identification of putative substrates for proteases can be an important step in assigning function to these proteases. In drug discovery, the assignment of function to genes is critical in determining whether these genes are likely to play a role in a

10 given disease process. If a protease is selected as a target for inhibition in a disease process, it will be necessary to devise an *in vitro* assay for the enzymatic activity of this protease. Information relating to consensus sequences can be used to synthesise such peptides.

Methods for determining the cleavage sites of proteases by selecting substrates from peptide libraries currently rely upon *in vitro* techniques. One such method employs phage

15 display of peptide libraries (Matthews and Wells, 1993). A library of peptides is constructed and expressed on the surface of bacteriophages so that if cleavage occurs at one of these sequences then they are not retained upon an affinity column. Determination of the DNA sequence of the gene encoding the peptides which are not retained on the column allows deduction of the cleavage sites for proteases. This method was employed in the analysis of

20 the specificity of stromelysin and matrilysin (Smith *et al.*, 1995). A second approach uses the activity of proteases against mixtures of substrates to derive information about protease cleavage sites (see for example Berman *et al.*, 1992; Petithory *et al.*, 1991). A version of this approach, positional scanning of synthetic combinatorial libraries (Pinilla *et al.*, 1992), was applied to the interleukin-1β converting enzyme to reveal a novel consensus sequence (Rano

25 *et al.*, 1997). An aldehyde derivative of this novel consensus sequence was a potent inhibitor of the enzyme. Subsequently the same approach has been applied to members of the Caspase family (Thornberry *et al.*, 1997), allowing the deduction of consensus cleavage sites for members of this family and some conclusions to be drawn about the biological role of these proteases.

30   Both of the methods described above are performed *in vitro* and require samples of purified protease. In addition, the second method is not easily applicable to protease

- 8 -

substrates which are more than four amino acids long, since the complexity of the mixtures

employed becomes too great. The heterologous cells described in this invention can be used

to determine the sequence specificity of the protease. Instead of providing a transcription

factor precursor with a known substrate for the protease, a library of such precursors is

5 constructed such that each precursor contains a different sequence between the membrane

anchor and the transcription factor. This library is then introduced into cells which contain

the protease and the target gene. Cells in which the target gene becomes activated will

therefore have expressed a precursor which contains a cleavage site for the protease. Upon

recovery of the gene encoding the precursor from these cells, the sequence of the cleavage site

10 may be deduced from the sequence of the DNA which encodes it. Combination of the results

of several such screens will allow a consensus sequence to be deduced. In comparison to the

methods described above, this system is an intracellular system, so that no purification or

expression of any protease is required; only a DNA sequence encoding the protease is

required. This allows an assessment of the activity of a protease in a cellular environment.

15   Therefore in a further aspect of the invention we provide a method for identifying a

substrate peptide sequence of a protease, which method comprises:

      (i)   creating a number of differing gene constructs which code for a transcription

      factor precursor, which comprises a transcription factor linked to a membrane

20   anchoring domain via a putative protease cleavage site wherein different putative

      protease cleavage sites are coded within the different gene constructs;

      (ii)   introducing the gene constructs into cells which contain the protease and a

      target gene under the control of the transcription factor; and

25

      (iii)   in each cell detecting whether the protease has cleaved the putative protease

      cleavage site to release the transcription factor by measuring expression of the target

      gene.

30   Where the sequence of the putative protease cleavage site introduced into each cell is

not known, for example where the sequences are generated by random mutagenesis or a

- 9 -

peptide library is generated by combinatorial techniques, then the following additional steps
are required in order to elucidate the sequence.

    (iv)   recovering the gene construct from cells with measurably increased levels of
5   target gene expression; and

    (v)   determining the sequence of each protease cleavage site recovered.

        A further aspect of the invention relates to methods for altering the specificity of
10 proteases. Many proteases display specificity for the amino acid sequence surrounding the
peptide bond which they cleave. If this sequence specificity could be altered then it might be
possible to design proteases which would cleave target proteins at desired sites. Such reagents
could be useful therapeutic agents. For example they could be used to cleave viral proteins
and so prevent a viral infection. Attempts to alter protease specificity have had some limited
15 success. In general, two approaches have been taken to the alteration of the substrate
specificity of proteases (Leis and Cameron, 1994). In the first approach, the three
dimensional structure of the protease is known and is used to predict the effects that amino
acid side chain alterations will have upon substrate recognition. This rational approach to
specificity alteration was first used to show that changes to side chains at the active site of
20 trypsin altered the substrate preferences of this enzyme (Craik *et al.*, 1985). A successful
alteration of specificity was achieved by Khouri *et al.* (1991), who replaced two amino acids
from papain with corresponding residues from cathepsin B and were effectively able to
change the specificity of the enzyme from that of papain to that of cathepsin B. In contrast,
however, when residue changes in chymotrypsin, predicted, on the basis of structural analysis,
25 to alter the specificity to that of trypsin, were made, no such change was seen (Venekei *et al.*,
1996). Evidently, residues other than those which directly contact the substrate are important
in determining the specificity of some proteases. Since rational alteration of protease
specificity has enjoyed only limited success, the second approach, that of random or semi-
random mutagenesis coupled with genetic selection techniques, has been pursued in other
30 cases. Mutants of a *Streptomyces* protease (Sidhu and Borgford, 1996) and *Lysobacter* alpha-
lytic hydolase (Graham *et al.*, 1993) with some degree of altered specificity have been

obtained using methods in which proteases are secreted from bacterial colonies and the

formation of clear plaques around a colony is measured.

The effect on the specificity of a protease following modification of the primary

peptide sequence may be measured using the cells described in this invention. Imagine that a

5 protease will cleave sequence 1, but not sequence 2. Sequence 1 may be introduced into a

precursor protein to create precursor 1. In cells which contain precursor 1 and the protease, the

target gene is activated. Sequence 2 may be introduced into the precursor protein to create

precursor 2. In cells which contain precursor 2 and the protease, the target gene is not

expressed. The gene encoding the protease is then subjected to some form of mutagenesis and

10 the mutated protease genes are introduced into cells containing precursor 2 and the target

gene. In those cells where the target gene is activated, the protease is now capable of cleaving

precursor 2. The genes encoding the proteases may be recovered from these cells and the

sequence of the new proteases deduced from the DNA sequences of the genes. New proteases

obtained in this manner may be able to recognise sequence 2 but not sequence 1, in which

15 case we may speak of the specificity of the proteases having been altered. Alternatively they

may be able to cleave both sequence 1 and 2, in which case we may described their specificity

as having been relaxed.

The reverse procedure may also be carried out if it is desirable to obtain a protease

which does not recognise a certain site. A sequence which the protease does cleave is

20 introduced into the precursor protein. In the presence of the protease, the target gene is

activated. The gene encoding the protease is then subjected to some form of mutagenesis and

the mutated protease genes are introduced into the cell containing the precursor and the target

gene. In those cells where the target gene is not activated or displays reduced activation, the

protease is now either incapable of cleaving the precursor protein, or has a reduced ability to

25 cleave respectively. The genes encoding the proteases may be recovered from these cells and

the sequence of the new protease deduced from the DNA sequence of the gene. In this way it

may be possible to "restrict" the specificity of a protease. If a protease recognises several

substrates (i.e. has a broad substrate range) it may be possible to restrict the specificity of the

protease so that it only recognises one substrate. This is useful in designing therapeutics that

30 will have few side effects.

Therefore in a further aspect of the invention we provide a method for altering the specificity of a protease which method comprises:

(i)   creating a protease gene constructs with a different coding variation from the
5   wild type protease peptide sequence;

(ii)   introducing the protease gene construct into a cell containing a target gene under the control of a transcription factor and a transcription factor precursor which comprises the transcription factor linked to a membrane anchoring domain via a
10   protease cleavage site; and

(iii)   detecting in each cell whether the altered protease has cleaved the protease cleavage site to release the transcription factor by measuring expression of the target gene.
15

Where the sequence of the protease introduced into each cell is not known, for example where the sequences are generated by random mutagenesis or a peptide library is generated by combinatorial techniques or gene shuffling techniques are used, then the following additional steps are required in order to elucidate the sequence.
20

(iv)   recovering the protease gene from the cells with measurably increased levels of target gene expression; and

(v)   determining the sequence of the recovered protease genes.
25

Further aspects of the invention comprise the individual constructs, plasmids and yeast strains disclosed in the accompanying Tables, Examples and Figures.

For all of the above aspects of the invention it is preferred that the protease cleavage site and the membrane binding domain are derived from different proteins.

30   The invention will now be illustrated but not limited by reference to the following Tables, Examples and Figures wherein:

- 12 -

Table 1 lists the plasmids used in Example 1. The first column gives the plasmid reference number. The second column ("Marker") indicates which selectable marker the plasmids contain. The third and fourth column describe the expression cassette which is located within each plasmid. The third column gives the promoter region which is used

5 (either *ACT1* (actin gene promoter) or *ADH1* (alcohol dehydrogenase gene promoter) from *S. cerevisiae*). The fourth column describes the coding region which is under the control of the promoter. The following conventions are used: "Gal" is the DNA binding domain of Gal4; "mVP" is an activation domain; "HMG" is the amino terminal region of the Hmg1 protein; "tev" is the wild type cleavage site for TEV protease; "TEV protease" comprises a methionine

10 fused to amino acids 2051-2279 of the tobacco etch virus polyprotein.

Table 2 displays the activity of the lacZ reporter gene in the *S. cerevisiae* strain NLY2::185 transformed with different combinations of plasmids (see Example 1). The first column indicates the number of the transformation. The second column indicates which

15 proteins are being expressed within the cells. The third column indicates which plasmids have been transformed into the NLY2::185 cells. The fourth column indicates the supplements omitted from the medium (thus UHL- indicates that uracil, histidine and leucine were omitted). The final two columns indicate the reporter gene activity as scored by the blue colour on X-Gal indicator plates (fifth column) and as measured in liquid culture in mOD/min

20 per OD of culture (sixth column). For transformations which displayed a low reporter gene activity (1, 4, 5, 6 and strain alone) six colonies were grown to mid-log phase and assayed. For the three transformations which displayed a high reporter gene activity (2, 3, and 7) fifteen colonies were grown to mid-log phase and assayed. For each set of reporter gene measurements the average and standard deviation are shown.

25

Table 3 shows the results of an alanine scan performed on the TEV protease cleavage site (see Example 2). Plasmids encoding the wild type and variant TEV protease cleavage site were cotransformed into NLY2::185 with a plasmid encoding the TEV protease (LDD208) and selected on UHL- media. Cleavage of the wild type site occurs between the Q and S

30 residues of the sequence ENLYFQS. Q is designated position -1 and S as position +1. Changes are denoted in the following way: x(y)z, where x is the parental amino acid, y is the

position of this amino acid in the cleavage site, and z is the residue to which it has been

changed. Thus Q(-1)A denotes that the Q at position -1 is changed to A. All plasmids are

derivatives of LDD882 (Table 1) and thus contain the *LEU2* marker and the *ACT1* promoter.

The first column indicates the transformation number. The second column gives the plasmid

5   numbers and the third column the precursor proteins which are expressed from these

plasmids. The fourth column displays the sequence of the cleavage site with the alanine

marked. The fifth column shows the activity of the lacZ reporter gene as scored used

indicator plates. The sixth column shows the activity of the reporter gene as measured using a

liquid assay. For each member of the alanine scan, approximately 20 colonies were taken in a

10   single loop. The resulting culture was assayed three times; the final column shows the mean

and standard deviation of these measurements.

Table 4 shows different ways in which the specificity profile of a protease can be

changed. The "profile" of the parental protease may be characterised against a variety of

related cleavage sites. For example the parental protease may be capable of cleaving a wild

15   type cleavage site and variant 1, but not variant 2 (column 2); this profile of activity is the

"wild type" profile. If a derivative of the protease retains the ability to cleave at the wild type

site and variant 1 but can also cleave variant 2, then the specificity of the protease in this

derivative is "relaxed" (column 3). Conversely, if a derivative of the protease can cleave at

the wild type site but cannot cleave either variant 1 or 2, then the specificity of the protease in

20   this derivative is "restricted" (column 4). Protease derivatives may also be obtained which

have lost the ability to cleave the wild type cleavage site but have gained the ability to cleave

at a site not recognised by the parental protease (variant 2). In this case the specificity of the

protease has been "altered" (column 4).

25   Table 5 displays the effects of deletions of the TEV protease upon the specficity of the

protease. Plasmids encoding proteases were transformed with the cleavage site variants

indicated in the first column. The results of the transformation were scored on X-Gal

indicator plates and were also measured using the liquid β-galactosidase assay. For each

transformation, approximately 20 colonies were taken in a single loop. The resulting culture

30   was assayed three times; the final column shows the mean and standard deviation of these

measurements.

Figure 1 is a diagram illustrating the design of a protease-dependent gene switch. The cell on the left contains: a membrane (1); a transcription factor fusion (2-5) comprising a membrane anchoring domain (2), a protease cleavage site (3), a DNA binding domain (4) and a transcription activation domain (5); and a target gene comprising a promoter containing one

5 or more binding sites for the DNA binding domain (6) and the transcribed region of the gene (7). In this cell the target gene is switched off since the transcription factor is membrane anchored and is unable to activate transcription. In the cell on the right a protease (8) which can act at the protease cleavage site has been expressed. Cleavage of the transcription factor fusion results in release of a transcription factor which can bind the promoter of the target

10 gene and activate target gene expression (indicated by the arrow).

Figure 2 shows the structures of the transcription activator fusions used in Example 1 (see also Table 1). Plasmid LDD883 encodes the activator Gal-mVP. The coding region comprises a methionine linked to amino acids 2-147 of *S. cerevisiae* Gal4 (the DNA binding

15 region of Gal4 (abbreviated to "Gal" here) linked to a 28 amino acid peptide which contains two repeats of the sequence LDDFDLDMLG. This 28 amino acid region is the activation region referred to as minimal VP16 (abbreviated to "mVP"). The fusion of the Gal4 DNA binding domain to the activation domain is illustrated in the box at the bottom of the figure and is referred to as "Activator" in the other portions of this figure. LDD1123 encodes the

20 protein tev-Gal-mVP, which comprises a peptide containing a TEV protease cleavage site (abbreviated to "tev") fused to the amino terminus of Gal-mVP. The TEV protease cleavage site is underlined. Plasmid LDD1117 encodes the protein HMG-Gal-mVP, which comprises the amino terminal region of the *S. cerevisiae* Hmg1 protein (abbreviated to "HMG") fused, via a 9 amino acid linker, to the amino terminus of Gal-mVP. The plasmid LDD882 encodes

25 the activator HMG-tev-Gal-mVP, which is identical to HMG-tev-Gal-mVP except for the addition of a TEV protease cleavage site (underlined) to the linker region separating HMG and Gal-mVP.

Figure 3 shows the structures of the tobacco etch virus (TEV) protease and the

30 deletions of this protease used in Examples 1-3. The numbers refer to the sequence of the polyprotein encoded by the tobacco etch virus. The protease referred to as "full length" in this

- 15 -

work comprises a methionine fused to amino acids 2051-2079 and is encoded by plasmid

LDD208. LDD855 encodes a version of TEV protease in which the C-terminal deletion 16

amino acids are deleted. This deletion is referred to as TEV protease[ΔC-16]. LDD859

encodes a version of TEV protease in which a region of 10 amino acids has been deleted.

5 This deletion is referred to as TEV protease [ΔI-10]

Example 1: a protease-dependent gene switch

Introduction

10

In this Example we describe the configuration of a gene switch as depicted in Figure 1.

As a model protease we have used the tobacco etch virus (TEV) protease. (Carrington and

Dougherty, 1987). This plant viral protease was selected because it possesses the following

features:

15

(1) Both the protease and its substrate are well defined. The protease has a molecular weight

of 49kdal and is encoded by an open reading frame of 430 amino acids. However, only the

amino terminal 229 residues are required for activity of the protease (Dougherty and

Carrington, 1988). The protease cleaves the viral polyprotein at several positions. By

20 analysis of these cleavage sites a consensus recognition site for the protease was identified as

E-x-hy-Y-x-Q-S/G (Carrington and Dougherty, 1988; single letter code: x denotes a

nonconserved residue, hy indicates a hydrophobic residue), where cleavage occurs between

the glutamine residue and the serine/glycine residue. The importance of the conserved

positions in the consensus sequence is supported by a series of experiments in which

25 mutagenesis of the cleavage site was combined with *in vitro* protease assays (Dougherty *et al.*,

1988).

(2) The cleavage site will be recognised and cleaved by the protease when introduced into

foreign proteins. TEV cleavage sites have been placed into transcription factors (Smith and

30 Kohorn, 1991; DasMahapatra *et al.*, 1992) and into the *E. coli* SecA protein (Mondigler and

Ehrmann, 1996).

- 16 -

(3) TEV protease does not appear to require activation from an inactive form, but rather is constitutively active. Therefore, no other components other than a gene encoding the protease open reading frame needs be introduced into a cell in order to express an active TEV protease.
(4) Despite being constitutively active, TEV protease is not toxic when expressed in yeast
5 cells (*S. cerevisiae*; Smith and Kohorn, 1991; DasMahapatra *et al.*, 1992) and in bacterial cells (*E. coli*: Marcos and Beachy, 1994, Mondigler and Ehrmann, 1996).

As a substrate for TEV protease we designed the following transcription factor precursor (as illustrated in Figure 1). As a membrane localisation domain, we selected the amino terminal region of the *S. cerevisiae* HMG-CoA reductase enzyme (Hmg1p). The
10 amino terminal region of this protein (amino acids 1-526) contains seven transmembrane spanning regions and is inserted into the membrane of the endoplasmic reticulum (ER) so that the carboxy terminus is exposed to the cytoplasmic face the ER (Basson *et al.*, 1988; Senstag *et al.*, 1990). A fusion of the green fluorescent protein to this region was concentrated in the perinuclear region (Hampton *et al.*, 1996b), suggesting that fusion protein located at the
15 carboxy terminus may project into the nuclear compartment. We made fusions at the carboxy terminus of the first 526 amino acids of Hmg1p. We would therefore expect the heterologous domains fused to the Hmg1p amino terminal region to be located immediately next to the ER membrane, on the cytoplasmic, and possibly nuclear, face of this membrane.

The transcription factor precursor comprises the HMG1p amino terminal domain
20 linked to a consensus cleavage site for TEV protease which in turn is linked to a transcription factor consisting of a DNA binding domain fused to a transcription activation domain. The DNA binding domain is derived from the *S. cerevisiae* protein Gal4 (Keegan *et al.*, 1986) and the activation domain is based on the herpes simplex virus VP16 transcription activation domain (Triezenberg *et al.*, 1988). As a host cell we used the budding yeast *S. cerevisiae*.
25 As a target gene for the transcription factor fusion we employed the *E. coli* lacZ gene, which has been used in a number of studies to measure the strength of Gal4-based transcription factors (for example Ma *et al.* 1987a;1987b).

30

Methods

Yeast strains

5   The experiments used in this study use the *S. cerevisiae* strain NLY2::185. To
construct this strain, the reporter plasmid JP185 (J. Pearlberg Ph. D. thesis, Harvard
University, 1994)) was cleaved within the *URA3* gene and integrated into NLY2 (Lehming *et
al.*, 1994: *α, ura3-52, his3-Δ200, leu2-3,112 , lys2-Δ(683-3543), trp1, ade2-101, gal4-542,
gal80-538*, obtained from Norbert Lehming, Harvard University) at the *ura3-52* locus,
10   reconstituting a functional URA3 gene.

Plasmids

        Plasmids in this and the following examples were constructed using fragments of
15   existing plasmids, fragments obtained from other plasmids by the polymerase chain reaction
and synthesised oligonucleotides. These were linked together by standard techniques
(Sambrook *et al.*, 1989). For each plasmid we describe the structure of the final plasmid,
rather than describing the steps by which the plasmid was made. It will be apparent to the
person of ordinary skill that there are many possible steps by which the plasmids we described
20   could be constructed. The structures of the plasmids are described by reference to sequences
in public databases. Sequences which link these fragments are derived from synthesised
oligonucleotides and we provide sequences of the linking region. (Note that since the
oligonucleotides used to incorporate such linking regions may have been synthesised with
additional sequences which were removed, perhaps by restriction enzyme digestion, during
25   construction of the final plasmid, the linking sequences do not necessarily correspond to the
oligonucleotides which would have been synthesised.)

        LDD208 is a plasmid from which a truncated version of the tobacco etch virus (TEV) protease
(part of the NIa gene) is expressed under the control of the *S. cerevisiae ADH1* promoter. The
30   backbone of this vector is the plasmid RS313 (Sikorski and Hieter, 1989). Between the *Sac I*
and *Kpn I* restriction enzyme sites in the polylinker of this vector we have placed an

- 18 -

expression cassette comprising the promoter of the *S. cerevisiae ADH1* gene, the coding

region of TEV protease and the transcription termination region region of the *S. cerevisiae*

ADH1 gene. The *ADH1* promoter sequence consists of a 1.4kb *Bam HI-Hind III* fragment

from the plasmid pADNS (Colicelli *et al..*, 1989). The *Hind III* site is followed immediately

5 by the linker:

GCGGCCGCCCACCATG

This linker introduces a translation initiation codon. This codon is linked in frame to a

10 sequence which encodes the active portion of the TEV protease. This sequence corresponds

to nucleotides 6295-6981 of Genbank entry g335201 (Allison *et al.*, 1986). This encodes

amino acids 2051-2279 of the tobacco etch virus polyprotein precursor, or amino acids

202-430 of the predicted 430 amino acid Nla protein released by proteolysis from the

polyprotein precursor. This sequence is followed by a linker which incorporates an in-frame

15 stop codon and an *Eco RI* site:

TAAGAATTC

Following the *Eco RI* site is a sequence which contains the transcriptional termination

20 region of the *S. cerevisiae ADH1* gene. This fragment is the 0.6kb *Eco RI-Bam HI* fragment

from the plasmid pADNS (Colicelli *et al.*, 1989). The protein encoded by LDD208 is

illustrated in Figure 3.

LDD882 is a plasmid from which the following fusion protein is expressed: a seven

25 transmembrane spanning region of *S. cerevisiae* HMG-CoA reductase linked to a cleavage

site for TEV protease linked to the DNA binding domain of the *S. cerevisiae* Gal4 protein

linked to a synthetic transcription activation domain based upon the herpes simplex virus

VP16 activation domain. The backbone of this vector is the plasmid RS315 (Sikorski and

Hieter, 1989). Between the *Sac I* and *Kpn I* restriction enzyme sites in the polylinker of this

30 vector we have placed an expression cassette comprising the promoter of the *S. cerevisiae*

ACT1 gene, the coding region of the fusion protein outlined above and the transcription

termination region region of the *S. cerevisiae ADH1* gene. The promoter region of *ACT1* consists of a 0.65kb DNA fragment obtained by polymerase chain reaction from *S. cerevisiae* genomic DNA. The fragment corresponds to nucleotides 8-671 of Genbank entry g170985 and is flanked at the 5' end by a Sac I site and at the 3' end by a *Hind III* site. Following the

5 *Hind III* site is a region of the *HMG1* gene from *S. cerevisiae* which encodes amino acids 1-526 of *S. cerevisiae* HMG-CoA reductase (Basson *et al.*, 1988). This sequence corresponds to nucleotides 112-1698 of Genbank entry g171685. Following this is the linker:

CTGCAGACTAGTACTGAAAATTTGTACTTCCAATCTGGTACCCATGGT

10

This linker encodes the peptide LQTSTENLYFQSGTHG. The sequence ENLYFQS is a cleavage site for the TEV protease (Dougherty *et al.*,1988). This sequence is followed by a sequence encoding the DNA binding domain (amino acids 2-147) of *S. cerevisiae* Gal4. The sequence corresponds to nucleotides 427-864 of Genbank entry g171557 (Laughon and

15 Gesteland 1984). Following this sequence is the linker:

CCCCGGGTCGAGTTGGATGACTTCGACTTAGATATGTTGGGTTTGGATGACTTCG
ACTTAGATATGTTGGGTGTCGACACTAGTTAACTAGCGGCCGC

20   This linker encodes the peptide PRVELDDFDLDMLGLDDFDLDMLGVDTS. This protein fragment contains two copies of a peptide (LDDFDLDMLG) based on the amino acid sequence of part of the activation region of the herpes simplex virus VP16 protein (amino acids 440-449 of the protein encoded by Genbank entry g330318). Following this peptide is a stop codon and a *Not I* site. The *Not I* site is linked to a DNA fragment which includes the

25 transcription terminator region of the *S. cerevisiae ADH1* gene. This fragment is the 0.6kb *Not I-Bam HI* from the plasmid pADNS (Colicelli *et al.*, 1989).

LDD1117 encodes a plasmid which is identical to LDD882, except that the linker region between the Hmg1 fragment and the Gal4 fragment is the sequence:

30

CTGCAGACTAGTACTGGTACCCATGGT

- 20 -

This encodes the amino acid sequence LQTSTGTHG.

LDD883 is a plasmid from which a fusion protein comprising the DNA binding domain of
Gal4 (amino acids 1-147) fused to the minimal activation domain based upon VP16 is
5   expressed under the control of the *S. cerevisiae ACT1* promoter. This plasmid is similar to
LDD882 except that the sequences which encode the HMG-CoA reductase region and the
TEV protease cleavage site are replaced by a sequence which supplies an initiation codon for
Gal4:

10 GAAGCAAGCCTCCTGAAAGATG

LDD1123 is a plasmid from which a fusion protein comprising a TEV protease cleavage site
linked to the DNA binding domain of Gal4 and the minimal activation domain based on VP16
is expressed. This plasmid is similar to LDD882 except that the sequences which encode the
15 HMG-CoA reductase is replaced by an initiation codon.

Yeast methods

Yeast was manipulated according to standard protocols (Sherman, 1991; Ausabel *et
20 al.*, 1993). Yeast cultures were grown in synthetic complete media lacking the appropriate
nutrients to allow for selection of transformed plasmids. This media was supplemented with
2% (w/v) glucose as a carbon source. Plasmids were transformed in yeast using the lithium
acetate procedure (Ito *et al.* 1983). 3 days after transformation, colonies were transferred to
X-Gal indicator plates (Ausabel *et al.*, 1993), and 24 hours later were scored for the degree of
25 blue colour. Liquid β -galactosidase assays were performed in microtitre plates using a
modification of the method of Dixon *et al.* (1997). A final concentration of 5mM substrate
(chlorophenol red galactopyranoside (CPRG) -approximately 10xKm) was used in the
reaction. Single colonies or a loop of approximately 20 colonies was inoculated into 10ml
miminal medium plus supplements. This culture was grown for approximately 40 hours. The
30 cultures were diluted back to an OD600 of 0.1. 120μl of each culture was transferred to wells
of a Costar 96-well microtitre plate. 30μl of a 5x reaction cocktail was added and an initial

absorbance at 570nm read. The initial rate of reaction (production of colour at 570nm) was
then measured over a ten minute period after addition of substrate.

A unit is defined as the rate of production of CPRG (mOD570/min) divided by the
optical density of the culture in the microtitre plate. A culture with an optical density of 0.1 at
5 600nm (measured in spectrophotometer with a 1cm path length) has an optical density of
0.0116 in Costar 96-well microtitre plates in a Molecular Devices platereader. Since the
culture is diluted by a factor of 1.25 by the addition of reaction cocktail, units are calculated
as:

10 Units = (rate of OD570 change) x 1.25/0.0166

Results and Discussion

Structure of a chimaeric transcription factor
15

The components of the chimaeric transcription factor constructed have been
previously described, with the exception of the transcriptional activation domain. The
activation domain of the herpes simplex virus VP16 protein is acidic (Treizenberg *et al.*,
1988) and very potent (Sadowski *et al.*, 1988). The negative charges provided by the acidic
20 residues are required for the transcription function (Cress and Triezenberg,1991), as is a
critical phenylalanine residue (F442: Regier *et al.*, 1993). A property of the VP16 activation
domain, and other strong acidic activation domains, such as that from RelA (Blair *et al.*,
1994) is the ability of small regions from the activation domain to provide strong activation
domains when multimerised. Seipel *et al.* (1992) showed that an 11 amino acid sequence
25 from the VP16 activation domain (amino acids 437-447: DALDDFDLDML) including the
critical phenylalanine residue, though weak activator by itself, is as strong as native VP16 as a
tandem dimer. This experiment was performed in mammalian cells. We find that an
overlapping 10 amino acid stretch derived from this region of VP16 (amino acids 439-448:
LDDFDLDMLG), is a weak activator in the yeast *S. cerevisiae*, but when dimerised is a very
30 strong activator in both *S. cerevisiae* and mammalian cells (unpublished results). The dimeric

- 22 -

form of this peptide (two tandem copies of amino acids 439-448) is the activator we use in the experiments described here. We refer to this activator as minimal VP16 (mVP).

Construction of a genetic system for TEV protease

5

    To demonstrate that we could configure a cell in which reporter gene activity is dependent upon both the presence of a protease and the presence of a cleavage site for the protease, we performed the experiment summarised in Table 2. The plasmids used in this experiment are summarised in Table 1, and the structures of the proteins expressed from these 10 plasmids are illustrated in Figures 2 and 3. The experiment consists of the transformation of expression plasmids into the reporter strain NLY2::185. This strain contains a reporter gene (*E. coli* β-galactosidase) which is under the control of two binding sites for Gal4. Since NLY2::185 contains no endogenous Gal4, the reporter gene activity reflects the presence of Gal4-based transcription factors expressed from plasmids. We transformed plasmids into 15 NLY2::185 and measured reporter gene activity by assessing the extent of blue colour of colonies after 24 hours on indicator plates (scored from 0 for no blue colour to 4 for very blue) and by measuring reporter gene activity in liquid culture. For reporter gene assays we measured the enzyme activity in cultures grown from six different colonies for transformations which displayed low reporter gene activity, or fifteen different colonies for 20 transformations which displayed a high reporter gene activity.

    In the host strain NLY2::185 (i.e. in the absence of Gal4-based transcription factors) the reporter gene is essentially inactive (see Table 1). A set of transformations were performed using plasmids containing transcription activators based on Gal-mVP. In discussing the results obtained below we make the assumption that the activator proteins are 25 expressed to similar levels. This assumption seems reasonable since it is known that both Hmg1 and Gal4 are relatively stable, and all the activator proteins were expressed from the same promoter (*ACT1*). Gal-mVP is able to activate the reporter gene to a very high level (13061 units: transformation 2). Fusion of the cleavage site for TEV protease to the amino terminus of Gal-mVP to create tev-Gal-mVP had little effect on the ability of Gal-mVP to 30 activate transcription (10431 units, transformation 3), showing that the TEV protease cleavage site does not, by itself, have any inhibitory effect on Gal-mVP. By contrast, fusion of the

amino terminal domain of HMG-CoA reductase to Gal-mVP had a dramatic effect, completely inhibiting the ability of Gal-mVP to activate the reporter gene. This was observed with a protein that contained a TEV protease cleavage site (HMG-tev-Gal-mVP: 11 units-transformation 5) and with a protein which did not contain a TEV protease cleavage site

5 (HMG-Gal-mVP: 5 units - transformation 4). Thus the membrane-spanning portion of HMG-CoA reductase is very effective in the inhibition of the activity of fused transcription factors.

By itself, TEV protease was unable to activate reporter gene transcription (1 unit: transformation 1). Cotransformation of the plasmid encoding TEV protease with the plasmid encoding the substrate protein lacking a TEV protease cleavage site (HMG-Gal-mVP) did not

10 result in activation of the reporter gene, indicating that there were no sequences within HMG-Gal-mVP capable of acting as substrates for TEV protease (6 units: transformation 6). By contrast, cotransformation of the plasmid encoding TEV protease with the plasmid encoding the substrate protein containing a TEV protease cleavage site (HMG-tev-Gal-mVP) resulted in a strong activation of reporter gene output (8506 units: transformation 7). The reporter gene

15 activity in this experiment was of a similar level to that observed with the activators Gal-mVP and tev-Gal-mVP. Although the activity observed with Gal-mVP appears higher than that obtained with tev-Gal-mVP, the variation in this activity was very high. We also observed that cultures of NLY2::185 containing Gal-mVP grew very slowly. This could be attributable to a nonspecific inhibition of gene transcription by the strong activation domain mVP

20 ("squelching": see Gill and Ptashne (1988)). However, cultures of NLY2::185 containing tev-Gal-mVP (which possesses the same transcription activation domain) or the combination of HMG-tev-Gal-mVP and TEV protease grew as well as the host strain itself. For the purposes of this experiment the comparison of the reporter gene activity obtained with HMG-tev-Gal-mVP and TEV protease (transformation 7) with that obtained with tev-Gal-mVP

25 (transformation 3) shows that TEV protease expression restores reporter gene expression fully.

These results demonstrate that (1) the HMG domain inhibits transcription factor function whereas the protease cleavage site does not, and (2) using a transcription factor inhibited by fusion to the HMG domain, reporter gene activation can be obtained by a

30 protease, but only if the protease cleavage site is present within the transcription factor fusion. Thus in the cell generated in transformation 7, the equivalent of the cell depicted on the right

- 24 -

hand side of Figure 1, reporter gene activity is dependent both upon the presence of the protease and the presence of a protease cleavage site. The utility of this system to study the protease and the cleavage site will be enlarged upon in Examples 2 and 3. We note here that the system provides a very simple output for protease activity (blue colour on X-Gal indicator

5 plates), and that using yeast as a host cell provides the ability to screen many thousands of colonies simultaneously.

The true level of stimulation by TEV protease expression is difficult to measure because the activity of the reporter gene in the presence of the precursor HMG-tev-Gal-mVP is very low. Comparison of transformations 5 and 7 indicates that the level of reporter gene

10 stimulation by the protease is at least 700 fold. The level of reporter gene stimulation and the absolute level of target gene expression in the presence of protease may be manipulated according to the choice of components used in the system. Thus the use of weaker transcription activation domains in place of mVP will result in a lower absolute level of target gene expression. Alteration of the structure of the promoter controlling the target gene will

15 also affect the levels of expression that may be achieved. In general the inclusion of more binding sites for the released transcription factor, and the placing of these sites closer to TATA box, will result in higher levels of protease-induced target gene transcription (see Ptashne (1988) for parameters which affect transcription in eukaryotes).

The incorporation of such a system into a cell type allows the regulation of target

20 genes by the expression of a protease. A transcription factor fusion would be expressed in a cell. A target gene containing binding sites for the transcription factor would also be introduced into the cell. By expression of the protease capable of cleaving the transcription factor precursor, expression of the target gene would be obtained. One way to use such a system would be to regulate desirable genes; the system may also be used as a method of

25 amplifying another gene switch. The first switch would turn on the gene encoding the protease, which would then activate the target gene of the transcription factor.

Many viruses, including plant pathogens such as the tobacco etch virus, from which the TEV protease comes, and retroviruses, such as the HIV virus, are dependent upon the activity of a virally encoded protease for viral propagation. Two genes could be incorporated

30 into the cells of an organism. One gene would express a transcription factor fusion as described in this example, but having the TEV protease site replaced by a cleavage site for the

- 25 -

virally encoded protease. The other gene would be a toxic gene under the control of a
promoter containing binding sites for the transcription factor. The cells of the organism
would not be affected by the presence of these genes. However, if the virus were to infect any
of these cells, the expression of the virally-encoded protease would result in cleavage of the
5 transcription factor precursor, release of transcription factor, activation of the toxic gene and
consequent cell death. in this way it may be possible to kill the cell before mature virions are
produced from infection and stop the virus from propagating itself.

The system disclosed here provides a method to clone genes encoding proteases. If
the cleavage site for a protease is known, but the protease itself is not known, a cell may
10 configured in which a transcription factor precursor containing the cleavage site is expressed,
together with a reporter gene as a target. A library of DNA sequences would be placed in an
expression vector and this library transformed or transfected into the cell type. Cells which
show activation of the reporter gene may contain a gene encoding a protease which acts on the
cleavage site. This gene may be recovered and sequenced in order to deduce the sequence of
15 the protease. The library of sequences may consist of cDNA sequences or of genomic
fragments. Optionally the library would be constructed in such a way as to favour the
intracellular expression of encoded proteins.

The system described in this example may be used to screen for modulators of
proteases by the identification of compounds which affect the level of reporter gene induction.
20 A particular advantage of yeast cells is the ease with which encoded libraries may be
screened. Thus DNA encoding a library of peptides could be introduced into cells containing a
system in which a protease is activating a reporter gene. Cells in which an increase or a
decrease in the reporter gene expression occurs would be identified and the modulating
peptide identified by recovery and sequencing of the gene which encodes the peptide. The
25 peptide libraries used could be essentially random, or they could be based upon partial
randomisation of known protease inhibitors.

30

Example 2: Identifying protease substrates

Introduction

5   The protease system described in Example 1 can be used to study protease specificity.
In this example we show that the specificity of the protease as assessed in this cell-based
system agrees with that observed *in vitro*. The TEV cleavage site used in Example 1 (referred
to as the "wild type" site) is a naturally occurring TEV protease site which also conforms to
the consensus site (see Table 1 of Dougherty *et al.*, 1988). One method to assess the
10   contribution of amino acid residues to an event such as a protein:protein interaction or enzyme
catalysis is to perform an alanine scan (Cunningham and Wells, 1989). In an alanine scan,
each amino acid in a region of interest is replaced with alanine, and the effect of this
substitution is assessed on the event being studied. Replacement of an amino acid with
alanine effectively removes the side-chain of that amino acid. We describe in this example an
15   alanine scan of the TEV protease cleavage site, and show that the results obtained are
consistent with the known consensus sequence for TEV protease

Methods

20   Yeast strains and methods are as described in Example 1

Plasmids

LDD882 (see Example 1) contains a naturally occurring cleavage site for TEV protease
25   (ENLYFQS). This is referred to as the "wild type" cleavage site. Cleavage occurs between
the glutamine and serine residues and we number residues within this site relative to the
cleavage point; thus Q is -1 and S is +1. Within this sequence are sites for the restriction
endonucleases *Spe I* and *Nco I*. These restriction sites are unique within the plasmid LDD882.
Plasmids encoding variants of the cleavage site which differed at a single residue from the
30   wild type site were constucted by ligating oligonucleotides into LDD882 cleaved with *Spe I*

and *Nco I*. The following variants were constructed by inserting oligonucleotides containing the triplet GCT (encoding alanine) at the appropriate position:

LDD887; glutamate at -6 changed to alanine

5

LDD896; asparagine at -5 changed to alanine

LDD888; leucine at -4 changed to alanine

10 LDD897; tyrosine at -3 changed to alanine

LDD1102: phenylalanine at -2 changed to alanine

LDD889; glutamine at -1 changed to alanine

15

LDD890; serine at position +1 changed to alanine

Results and Discussion

20   The results of the alanine scan are presented in Table 3. We find that some alanine substitutions have very little effect on the reporter gene output, whereas some have a dramatic effect on the reporter gene output. These effects were scored subjectively on a scale of 0 to 4 on the indicator plates (0 indicating no blue colour, 4 being the colour observed with TEV protease and the precursor encoded by LDD882). Some heterogeneity was observed in the

25 blue colour of colonies in each the transformations. The scores represent the average colour of colonies. The reporter gene measurements were made on cultures grown from mixtures of approximately twenty colonies, in order to average the effects of colony variation. The reporter gene activity measurements are roughly in accord with the scores assigned on indicator plates, suggesting that the scoring system is a valid way to assess reporter gene

30 activity. Some amino acid changes in the alanine scan in a cleavage site give rise to an intermediate reporter gene output. We interpret these intermediate effects as indicating the

- 28 -

partial cleavage of the precursor and release of an intermediate quantity of Gal4. Intermediate
effects therefore derive from (in this example) cleavage sites which are not as good substrates
as the wild type site. These intermediate effects are important since they allow a "weighting"
to be assigned to cleavage sites (see below).

5   Substitution of alanine at positions -6, -5, -2 and +1 has little effect. At positions -3
and -4 there is a significant reduction in reporter gene activity, and at -1 a complete loss of
reporter gene activity. The consensus cleavage site for TEV protease is E-x-hy-Y-x-Q-(S/G),
where x denotes any amino acid and hy denotes a hydrophobic amino acid. Replacement of
the glutamine by alanine abolishes reporter gene activity. This is consistent with the strong

10 conservation of glutamine at position -1 of the cleavage sites for TEV protease and for the
cleavage sites of related viruses. Positions -3 is preferentially tyrosine. When this amino acid
was changed to alanine an intermediate phenotype was observed (scored as 2 out of 4 by blue
colour, measured as 2926 units). This suggests that tyrosine contributes to the consensus site
but is not as important as the glutamine at position -1. Dougherty *et al.* (1988) also found an

15 intermediate effect when they changed the tyrosine to alanine in an *in vitro* reaction. Position
-4 is usually leucine, valine or isoleucine, and alteration of this amino acid to an alanine also
results in an intermediate phenotype. In contrast positions -2 (phenylalanine) and -5
(asparagine) are not conserved, and alteration of these residues to alanine has little effect on
the reporter gene activity, indicating that the enzyme is as active on these substrates as it is on

20 the wild type substrate. The glutamate residue at position -6 was predicted to be important
since it is highly conserved; however alteration of this residue to alanine has little effect. This
change was not made by Dougherty *et al.* (1988), so no comparison with an *in vitro* result can
not be made. Although position +1 is normally glycine or serine, Dougherty *et al.* (1988)
show that a number of different amino acids can be tolerated at this position (though they did

25 not examine alanine), so it is not surprising that an alanine substitution has no effect on the
reporter gene activity. In general (and particularly for positions -5 to -1) the results of the
alanine scan are as might be predicted from the known consensus sequence for the TEV
cleavage site. This suggests that the protease reporter system is a valid approach to the
analysis of protease specificity. We also note that the intermediate levels of reporter gene

30 activation obtained with some of the cleavage site point mutants suggest that if this system

were to be used as a gene switch, then one way to modulate the strength of the switch would be to vary the peptide sequence of the protease cleavage site.

The reporter gene system may be used to study protease cleavage sites in two ways. Firstly, as described here, a sequence known to contain a protease cleavage site can be used as

5 a starting point. Defined changes are then introduced into this site and the effects of these changes on reporter gene output are observed. There is no need to recover the sequences encoding the cleavage sites, because these are characterised before being used in the experiment. Secondly, a library of sequences may be created as part of precursor protein. The protease would then be used to select sequences from this library. In this approach it would be

10 necessary to recover the DNA encoding the cleavage site in order to deduce the amino acid sequence of the site. The DNA encoding the cleavage site sequences that have been identified by reporter gene activation could be recovered by amplification through the polymerase chain reaction, or through recovery of the plasmid by transformation of *E. coli* with extracts of the cells which gave a positive output.

15 To determine a cleavage site for an protease, a peptide library could be placed between the membrane anchor and the transcription factor, encoded by partially randomised oligonucleotides. This would be introduced into cells with a reporter gene and the protease of interest so that only one (or very few) members of the library would be present in each cell. Cells in which reporter gene output is stimulated would be identified. The extent of reporter

20 gene activation would be scored for these cells, and the cleavage site present in the precursor deduced from the DNA encoding the precursor. Note that it would be important to confirm that any reporter gene activation is protease-dependent, and not due to some endogenous activity which is independent of the protease of interest. The information about the cleavage sites and the score of these sites would then be combined to generate a consensus sequence for

25 the protease under study. Several methods could be employed to deduce this consensus. The simplest approach would involve two steps. Firstly the sequences obtained would be aligned. Secondly, in the most favourable alignment, a consensus would be deduced to which each individual sequence contributes according to the score of that sequence. Thus better cleavage sites (which are associated with a higher reporter gene activity) would comprise a larger

30 component of the consensus sequence. This consensus site could then be used to search protein sequence databases to identify proteins which may be substrates for the protease.

- 30 -

The identification of a consensus cleavage site can also be used in inhibitor design. The peptide identified as a cleavage site may be directly convertible into an inhibitor. For example, Rano *et al.* (1997) identified a cleavage site for interleukin-1β converting enzyme and found that an aldehyde version of this peptide was a very potent inhibitor of the enzyme.

5 Alternatively, the peptide may be a starting point for the design of a peptidomimetic (reviewed by Giannis and Rubsam, 1997).

In general many proteases will be able to function in the context of the system we have described. It will be necessary to remove sequences which target the protease to cellular compartments or the cell surface. It may also be necessary to remove inhibitory sequences

10 that keep proteases in an inactive "pro" form. For example, the metalloproteinase TNFα converting enzyme (TACE) is an enzyme which is extracellular, cell-surface anchored and activated from a pro form during passage through the secretory pathway (Black *et al.*, 1997, Moss *et al.*, 1977). To prevent entry into the secretory pathway, the signal sequence (amino acids 1-17) should be removed. Amino acids 19-214 constitute a "pro" region which may

15 optionally be removed. The catalytic domain (215-473) may be expressed alone, or with some of the other domains, including the disintegrin domain (474-572) and also the transmembrane domain (672-694).

Example 3: Altering protease specificity

20

Introduction

Proteases are not generally completely specific for one substrate. Instead they will cleave, with varying efficiency, a set of related sequences. In considering how protease

25 specificity may be altered we may imagine the scenarios outlined in Table 4. A "parental protease" will have a profile of activity. It will be active against a wild type cleavage site and perhaps one variant of this wild type sequence (variant cleavage site 1) but not against another variant of this wild type sequence (variant cleavage site 2). This profile can be described as a "wild type" profile. Through mutagenesis protocols it may be possible to select for a

30 derivative of this protease (derivative 1) which can now act on variant 2 as well as variant 1 and the wild type site. Effectively this protease can recognise additional substrates, and its

- 31 -

profile has been "relaxed". In contrast, it may be possible to select for a derivative (derivative 2) which can only act on the wild type cleavage site. Since this derivative has lost the ability to cleave at variant cleavage site 1, it is described as having a "restricted" specificity. Finally a protease derivative which has lost the ability to recognise the wild type cleavage site, but

5   which can now recognise a cleavage site which it was formerly unable to recognise may be described as having an "altered" specificity. If this derivative still retains activity towards variant cleavage site 1, then its profile will overlap with the parental protease, whereas if this protease does not show activity towards variant cleavage site 1, then the specificity will not overlap.

10   Since, in the genetic system described in Example 1, reporter gene activity is dependent upon the presence of both protease and substrate, the system can be used to generate and study protein mutants with changed specificity. In this example we demonstrate this effect by describing two deletion versions of the TEV protease, one of which has a relaxed specificity and the other of which has a tightened specificity.

15

Methods

Yeast strains and methods are as described in Example 1.

Plasmids

20   LDD208 (see Example 1) contains the coding region of the TEV protease (amino acids 2051 to 2279). Within this plasmid there is a unique Not I restriction endonuclease site before the translation start codon, a unique *Spe I* site encompassing amino acids 2165-2166, and a unique *Eco RI* site after the translation stop. Deletion variants of the TEV protease were constructed using the the polymerase chain reaction combined with replacement of regions

25   between these sites.

LDD859 contains a deletion of ten amino acids (amino acids 2155-2164) in the centre of the protease, adjacent to the region encoding the *Spe I* site. This deletion, referred to as TEV protease[ΔI-10] (for internal deletion of ten amino acids) , was constructed by amplifying

30   from LDD208 a region encoding amino acids 2051 to 2155 using the oligonucleotides:

- 32 -

CCAGGCGGCCGCCCACCATGATATCGAGCACCATTTGT

CATGACTAGTTTGGAAGTTGGTTGTCAC

5   The DNA fragment obtained was digested with *Not I* and *Spe I* and ligated into LDD208
    cleaved with *Not I* and *Spe I* to obtain LDD859.

    LDD855 contains a deletion of 16 amino acids (amino acids 2264- 2279) at the carboxy
    terminus of the TEV protease. This deletion, referred to as TEV protease [ΔC-16] (for
10 carboxy-terminal deletion of sixteen amino acids) was constructed by amplifying from
    LDD208 a region encoding amino acids 2051 to 2264 using the oligonucleotides:

    CCAGGCGGCCGCCCACCATGATATCGAGCACCATTTGT

15 CATGGAATTCTTACTGAAAAGGCTCTTCAGG

    The DNA fragment obtained was digested with *Not I* and *Eco RI* and ligated into LDD208
    cleaved with *Not I* and *Eco RI* to obtain LDD 855.

20 **Results and Discussion**

          A set of amino and carboxy terminal deletions of TEV protease was made to examine
    the sequence requirements of the protease for activation of the reporter gene. It was found
    that removal of ten or more amino acids from the amino terminus of the protease encoded by
25 plasmid LDD208 caused a complete loss of activity against the wild type cleavage site (data
    not shown), whereas sixteen amino acids could be removed from the carboxy terminus
    without loss of activity. An internal deletion of ten amino acids also retained activity against
    the wild type cleavage site. The carboxy terminal deletion (TEV protease[ΔC-16]) and the
    internal deletion (TEV protease[ΔI-10]) were tested against a variety of cleavage sites with
30 single amino acid changes. These deletions were found to have a dramatic affect on the
    specificity of the protease. In this example we illustrate this effect by using three of the single

amino acids changes from the alanine scan (Table 5). The full length protease strongly activates a reporter gene in the presence of a transcription factor precursor containing a wild type cleavage site. It is moderately active against the L(-4)A and Y(-3)A cleavage sites, and inactive on the Q(-1)A cleavage site. The C terminal deletion has a profile which is relaxed

5 compared to the wild type protease. It is active against the wild type site, almost as active against L(-4)A and Y(-3)A sites as against the wild type site, and, significantly active against the Q(-1)A site. This deletion version of the protease has gained the ability to act on the Q(-1)A site. In contrast, the internal deletion, compared to full length protease, has a restricted specificity. Although it is strongly active on the wild type site, it does not act on

10 any of the other three sites.

It is possible that some of the effects observed are due to different levels of expression or different stabilities of the proteases. However, the ability of the carboxy terminal deletion to cleave at Q(-1)A is so striking that it seems unlikely that would be due to a substantially higher level of expression of this deletion. Similarly it would be unlikely that the complete

15 loss of ability of the internal deletion to act on the point mutants would be attributable entirely to a lower level of expression. Regardless of whether expression levels contribute to the effects observed, it is certainly clear that, functionally, the specificity of the TEV protease has been altered. The internal deletion is, in this setting, a highly specific protease, whereas the carboxy terminal deletion is a protease with a wider substrate specificity. Such proteases

20 may in themselves have utility. If the system as depicted in Figure 1 is to be used as a gene switch, it may be desirable to use a protease which is very specific, so that there is a minimal chance that other proteins will be cleaved within the cell. In this setting, the protease with the restricted profile would be appropriate. The carboxy terminal deletion displays a relaxed profile. A protease with a relaxed profile may be of use in applications where it is desirable to

25 maintain the diversity of a set of proteins or peptides. We show here that the carboxy terminal deletion has relaxed specificity for changes in the -1 position of the TEV protease cleavage site. It also displays a relaxed specificity for changes at the +1 position (data not shown). The +1 position of the cleavage site is also the amino terminal residue of the released protein. Imagine that, rather than releasing a transcription factor by protease cleavage, a library of

30 random peptides is being released by cleavage with the protease. If a restricted specificity protease is used then only those peptides which have certain amino acids at the amino

- 34 -

terminus will be released. However, if a protease with a relaxed specificity for position +1 of

its cleavage site is employed, then the peptides which are released will have a greater diversity

of residues at the amino terminal position.

This example shows that it is possible to detect and characterise proteases with an

5   altered profile of activity. The proteases used in this example were selected for study on the

basis that they were active against the wild type cleavage site. The approach can be used to

design novel specificity profiles of proteases if the generation of protease variants is combined

with a selection system based on reporter gene output. A library of genes encoding protease

variants could be generated by a number of methods, including chemical mutagenesis of the

10 gene and DNA amplification strategies which introduce mutations or which allow mixing of

sequences from homologous genes (Enell, L.P. and Loeb, L.A. (1998) Nature Biotech. 16,

234-235.

This library may then be transformed into cells which contain substrate proteins which

contain a certain cleavage site. The proteases that can act on this cleavage site will switch on

15 the reporter gene, enabling the cells which contain these proteases to be identified and the

genes encoding the proteases to be recovered. A novel protease obtained in this manner can

then be tested against a panel of substrates to determine its specificity profile, in particular to

determine whether it retains the ability to act on the wild type substrate (in which case the

specificity will be "relaxed") or whether it has lost the ability to act on the wild type substrate

20 (in which case the specificity will be "altered"). By following a series of such mutagenesis

and selection steps, it would be possible to evolve the specificity of a protease. The directed

evolution of catalytic activities is useful where rational design approaches are limited

(Kuchner and Arnold, 1997)

- 35 -

References

Allison R., Johnson R.E., Dougherty W.G. (1986) Virology 154, 9-20.

5 Ausabel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (1993) Current protocols in molecular biology. Chapter 13 (Wiley).

Basson, M.E., Thorsness, M., Finer-Moore, J., Stroud, R.M. and Rine, J. (1988) Mol. Cell Biol. 8, 3797-3808.

10

Baum, E.Z, Beberbitz, G.A. and Gluzman, Y. (1990). Proc. Natl. Acad. Sci. U.S.A. 87, 10023-10027.

Berman, J., Green, M., Sugg, E., Anderegg, R., Millington, D.S., Norwood, D.L., McGeehan, 15 J. And Wiseman, J. (1992) J. Biol. Chem. 267, 1434-1437.

Black, R.A. et al.. (1997) Nature 385, 729-733

Blair, W., S., Bogerd, H.P., Madore, S.J. and Cullen, B.R. (1994) Mol. Cell Biol. 14, 7226-20 7234.

Blobel, C.P. (1997) Cell 90, 589-592.

Brent, R. and Ptashne, M. (1981) Proc. Natl. Acad. Sci. USA 78, 4204-4208.
25

Carrington, J.C. and Dougherty, W.G. (1987) J. Virol. 61, 2540-2548.

Carrington, J.C. and Dougherty, W.G. (1988) Proc. Natl. Acad. Sci. U.S.A. 85, 3391-3395.

30 Casadaban et al. (1983) Meth. Enzymol. (1983) 100, 293-308.

- 36 -

Chalfie, M., Tu, Y., Euskirschen, G., Ward, W.W. and Prasher, D.C. (1994) Science 263, 802-805.

Colicelli, J. Birchmeier, C., Michaeli, T., O'Neill, K., Riggs, M. and Wigler, M. (1989) Proc. Natl. Acad. Sci. USA 86, 3599-3603.

Craik, C.S., Largman, C., Fletcher, T., Roczniak, S., Barr, P.J., Fletterick, R. and Rutter, W.J. (1985). Science 228, 291-297.

Cress, W.D. and Triezenberg, S.J. (1991) Science 251, 87-90

Cunningham, B.C. and Wells, J.A. (1989) Science 244, 1081-1085.

DasMahapatra, B., DiDomenico, B., Dwyer, S., Ma, J., Sadowski, I. and Schwartz, J. (1992). Proc. natl. Acad. Sci. U.S.A. 89, 4159-4162.

De Wet, J.R., Wood, K.V., Helinski, D.R. and DeLuca, M. (1985) Proc. Natl. Acad. Sci. USA 82, 7870-7873.

Dixon, G., Scanlon, D., Cooper, S. and Broad, P. (1997). J. Ster. Biochem. Mol. Biol. in press.

Doolittle, R.F. (1997) Nature 392, 339-342.

Dougherty, W.G. and Carrington, J.C. (1988) Ann. Rev. Phytopathol. 26, 123-143

Dougherty, W.G. and Semler, B.L. (1993) Microbiol. Rev. 57, 781-822.

Dougherty, W.G., Carrington, J.C., Cary, S.M. and Parks, T.D. (1988) EMBO J. 7, 1281-1287.

Fields, S. and Song, O. (1989) Nature 340, 245-246.

- 37 -

Giannis, A. and Rubsam, F. (1997) Adv. Drug Res. 29, 1-78.

Gill, G. and Ptashne, M. (1988) Nature 334, 721-724.

5   Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F.,
    Hoheisel, J.D., Jacq, C., Johnston, M., Louis, E.J., Mewes, H.W., Murakami, Y., Philippsen,
    P., Tettelin, H. and Oliver, S.G. (1996) Science 274, 562-567.

    Gorman, C., Howard, B. And Reeves, R. (1983) Nuc. Acids. Res. 11, 6731.
10
    Graham, L.D., Haggett, K.D., Jennings, P.A., Le Broque, D.S. and Whittaker, R.G. (1993)
    Biochemistry 32, 6250-6258.

    Hampton, R., Dimster-Denk, D. and Rine, J. (1996a) Trends in Biochem. Sci. 21, 140-145
15
    Hampton, R.Y, Koning, A., Wright, R. and Rine, J. (1996b) Proc. Natl. Acad. Sci. U.S.A. 93,
    828-833

    Hancock, J.F., Cadwallader, K., Paterson, H. and Marshall, C.J. (1991) EMBO J 10, 4033-
20 4039.

    Harrison, S.C. (1991) Nature 353, 715-719.

    Hirowatari, Y., Hijikata, M. and Shimotohno, K. (1995) Anal. Biochem. 225, 113-120.
25
    Ito, H., Fukada, K., Murata, K. and Kimura, A. (1983) J. Bacteriol 153, 163

    Keegan, L., Gill, G., and Ptashne, M. (1986) Science 231, 699-704.

30 Khouri, H.E., Vernet, T., Menard, R., Parlati, F., Laflamme, P., Tessier, D.C., Gour-Salin, B.,
    Thomas, D.Y. and Storer, A. (1991) Biochemistry 30, 8929-8936.

Kuchner, O. and Arnold, F.H. (1997) TibTech 15, 523-530.

Laughon, A. and Gesteland, R.F. (1984) Mol. Cell Biol. 4, 260-267.

5 Lehming, N., Thanos, D., Brickman, J., Ma, J., Maniatis, T. and Ptashne, M. (1994) Nature 371, 175-179.

Leis, J.P. and Cameron, C.E. (1994) Curr. Opin. Biotechnol. 5, 403-408.

10 Ma, J. and Ptashne, M. (1987a) Cell 48, 847-853.

Ma, J. and Ptashne, M. (1987b) Cell 51, 113-119.

Marcos, J.F. and Beachy, R.N. (1994) Plant Mol. Biol. 24, 495-503.

15

Materson, B.J. and Preston, R.A. (1994) Arch. Intern. Med. 154, 513-523.

Matthews, D.J. and Wells, J.A. (1993) Science 260, 1113-1117.

20 McCall, J.O., Kadam, S. and Katz, L. (1994) Bio/Technology 12, 1012-1016.

Mondigler, M. and Ehrmann, M. (1996) J. Bacteriol 178, 2986-2988.

Moss, M.L. and 25 others (1997) Nature 385, 733-736.

25

Pahl, H.L. and Baeuerle, P.A. (1996) Curr. Opin. Cell. Biol. 8, 340-347.

Pellman, D., Garber, E.A., Cross, F.R. and Hanafusa, H. (1985) Nature 314, 374-377.

30 Petithory, J.R., Masiarz, F.R., Kirsch, J.F., Santi, D.V. and Malcolm, B.A. (1991) Proc. Natl. Acad. Sci. USA 88, 11510-11514.

- 39 -

Pinilla, C., Appel, J.R., Blanc, P. and Houghten, R.A. (1992) Biotechniques 13, 901-905.

Ptashne, M. (1988) Nature 335, 683-689.

5 Rano, T.A., Timkey, T., Peterson, E.P., Rotonda, J., Nicholson, D.W., Becker, J.W.,
Chapman, K.T. and Thornberry, N.A. (1997) Chem. Biol. 4, 149-155.

Richman, D. (1996) Science 272, 1886-1888.

10 Sadowski, I., Ma, J., Triezenberg, S. and Ptashne, M. (1988) Nature 335, 563-564.

Sakai, J., Duncan, E.A., Rawson, R.B., Hua, X., Brown, M.S. and Goldstein, J.L. (1996) Cell
85, 1037-1046.

15 Sambrook J., Fritsch, E.F. and Maniatis, T. (1989)Molecular Cloning: A Laboratory Manual,
2nd Edition. Cold Spring Harbor Press

Seipel, K., Georgiev, O. and Schaffner, W. (1992) EMBO J. 11, 4961-4968.

20 Sengstag, C., Stirling, C., Schekman, R. and Rine, J. (1990) Mol. Cell Biol. 10, 672-680.

Sherman, F. (1991) Meth. Enzymol. 194, 3-21.

Shockett, P.E. and Schatz, D.G. (1996) Proc. Natl. Acad. Sci. U.S.A. 93, 5173-5176.
25

Sices, H.J. and Kristie, T.M. (1998) Proc. Natl. Acad. Sci. USA 95, 2828-2833.

Sidhu, S.S. and Borgford, T.J. (1996) J. Mol. Biol. 257, 233-245.

30 Sikorski, R.S. and Hieter, P. (1989) Genetics 122, 19-27.

Smith, T.A. and Kohorn, B.D. (1991) Proc. Natl. Acad. Sci. U.S.A. 88, 5159-5162.

- 40 -

Smith, M.M., Shi, L. and Navre, M. (1995) J. Biol. Chem. 270, 6440-6449.

Thornberry, N.A., Rano, T.A., Peterson, E.P., Rasper, D.M., Timkey, T., Garcia-Calvo, M., Houtzager, V.M., Nordstrom, P.A., Roy, S., Vaillancourt, J.P., Chapman, K.T. and Nicholson,
5 D.W. (1997) J. Biol. Chem. 272, 17907-17911.

Treizenberg, S.J., Kingsbury, R.C. and McKnight, S.L. (1988) Genes Dev. 2, 718-729.

Triezenberg, S.J. (1995) Curr. Opin. Genet. Dev. 5, 190-196.
10
Venekei, I., Szilagyi, L., Graf, L. and Rutter, W.J. (1996) FEBS Lett. 379, 143-147.

Wang, X., Sato, R., Brown, M.S., Hua, X. and Goldstein, J.L. (1994) Cell 77, 53-62.

15 Zlokarnik, G., Negulescu, P.A., Knapp, T.E., Mere, L., Burres, N., Feng, L., Whitney, M., Roemer, K. And Tsien, R. (1998) Science 279, 84-88.

Table 1

Plasmid number	Marker	Promoter	Protein
LDD208	HIS3	ADH1	TEV protease
LDD883	LEU2	ACT1	Gal-mVP
LDD1123	LEU2	ACT1	tev-Gal-mVP
LDD1117	LEU2	ACT1	HMG-Gal-mVP
LDD882	LEU2	ACT1	HMG-tev-Gal-mVP

Table 2

Transformation	Proteins	Plasmids	Selection	Score	Reporter Activity
(Strain alone)			U-	0	1 ±1
1	TEV protease	LDD208	UH-	0	1 ±1
2	Gal-mVP	LDD883	UL-	4	13061 ±4562
3	tev-Gal-mVP	LDD1123	UL-	4	10431 ±1084
4	HMG-Gal-mVP	LDD1117	UL-	0	5 ±3
5	HMG-tev-Gal-mVP	LDD882	UL-	0	11 ±3
6	TEV protease + HMG-Gal-mVP	LDD208 + LDD1123	UHL-	0	6 ±2
7	TEV protease + HMG-tev-Gal-mVP	LDD208 + LDD882	UHL-	4	8506 ±1920

Table 3

Transformation	Plasmid	Protein	Cleavage site	Score	Reporter activity
1	LDD882	HMG-tev-Gal-mVP	ENLYFQ S	4	8506 ± 1920
2	LDD887	HMG-tev[E(-6)A]-Gal-mVP	ANLYFQ S	3	8245 ± 277
3	LDD896	HMG-tev[N(-5)A]-Gal-mVP	EALYFQ S	4	8498 ± 502
4	LDD888	HMG-tev[L(-4)A]-Gal-mVP	ENAYFQ S	2	1665 ± 141
5	LDD897	HMG-tev[Y(-3)A]-Gal-mVP	ENLAFQ S	2	2926 ± 95
6	LDD1102	HMG-tev[F(-2)A]-Gal-mVP	ENLYAQ S	4	6620 ± 1187
7	LDD889	HMG-tev[Q(-1)A]-Gal-mVP	ENLYFA S	0	19 ± 10
8	LDD890	HMG-tev[S(+1)A]-Gal-mVP	ENLYFQ A	4	13314 ± 838

Table 4

	Parental protease	Protease		Protease derivative 3
		Protease derivative 1	Protease derivative 2	Protease derivative 3
Cleavage site				
Wild type cleavage site	+	+	+	-
Variant cleavage site 1	+	+	-	(+/-)
Variant cleavage site 2	-	+	-	+
Profile	Wild type	Relaxed	Restricted	Altered

Table 5

Cleavage site		Protease			
		TEV protease	TEV protease[ΔC-16]	TEV protease[ΔI-10]	
		LDD208	LDD855	LDD859	
Wild type site	Score	4	4	3	
LDD882	Reporter Activity	8506 ±1920	9483 ±692	9596 ±458	
L(-4)A	Score	2	3	0	
LDD888	Reporter Activity	1665 ±141	9601 ±555	24 ±6	
Y(-3)A	Score	2	3	0	
LDD897	Reporter Activity	2926 ±95	7547 ±353	11 ±10	
Q(-1)A	Score	0	1	0	
LDD889	Reporter Activity	19 ±10	2365 ±118	17 ±3	

CLAIMS

1.   A heterologous cell which comprises:

(i)   a transcription factor precursor which comprises a transcription factor linked to a membrane anchoring domain via a protease cleavage site, in which the membrane anchoring domain and protease cleavage site are not derived from the same protein;

(ii)   a protease which recognises the protease cleavage site in the transcription factor precursor; and

(iii)   a target gene under the control of the transcription factor, wherein if cleavage of the protease cleavage site by the protease is allowed to occur subsequent release of the transcription factor enhances expression of the target gene.

2.   A heterologous cell as claimed in claim 1 wherein the membrane anchoring domain is part of the N terminal sequence of the enzyme HMG-CoA reductase.

3.   A heterologous cell as claimed in claim 1 selected from a mammalian cell type or a yeast cell type.

4.   Use of a heterologous cell as claimed in claim 1 as a gene switch.

5.   Use of a heterologous cell as claimed in claim 1 to determine the substrate peptide sequence of a protease by determining whether there is expression of the target gene within a number of heterologous cells wherein within the number of heterologous cells a variety of different protease cleavage site sequences are coded between the transcription factor and membrane anchoring domain and therefore expression of the target gene indicates that the protease cleavage site is a substrate peptide sequence of the protease.

6.   Use of a heterologous cell as claimed in claim 1 to alter the specificity of a protease by determining whether there is expression of the target gene within a heterologous cell

containing a protease with a different peptide sequence than the normal protease and therefore
expression of the target gene indicates that the protease has successfully cleaved the cleavage
site.

5   7.   A gene switch mechanism comprising:

     (i)   a transcription factor linked to a membrane anchoring domain via a protease cleavage
             site;

     (ii)   a protease which recognises the protease cleavage site in the transcription factor
10   precursor; and

     (iii)   a gene placed under the control of the transcription factor, whereby enhanced
             expression of the gene occurs after cleavage of the protease cleavage site by the
             protease and thereby expression of the gene may be modulated by directly or indirectly
             affecting the activity or expression of the protease.

15

     8.   A gene switch amplification mechanism comprising:

     (i)   a transcription factor linked to a membrane anchoring domain via a protease cleavage
             site;

20   (ii)   a protease which recognises the protease cleavage site in the transcription factor
             precursor wherein expression of the protease gene is under the control of a regulatable
             gene switch; and

     (iii)   a gene placed under the control of the transcription factor, whereby enhanced
             expression of the gene occurs after expression of the protease gene is turned on by
25   cleavage of the protease cleavage site by the protease to release the gene transcription
             factor.

     9.   A method for identifying a substrate peptide sequence of a protease, which method
     comprises:

30

(i)   creating a number of differing gene constructs which code for a transcription factor precursor, which comprises a transcription factor linked to a membrane anchoring domain via a putative protease cleavage site wherein different putative protease cleavage sites are coded within the different gene constructs;

(ii)   introducing the gene construct into a cell which contain the protease and a target gene under the control of the transcription factor; and

(iii)   in each cell detecting whether the protease has cleaved the putative protease cleavage site to release the transcription factor by measuring expression of the target gene.

10.   A method for altering the specificity of a protease which method comprises:

(i)   creating a protease gene construct with a different coding variation from the wild type protease peptide sequence;

(ii)   introducing the protease gene construct into a cell containing a target gene under the control of a transcription factor and a transcription factor precursor which comprises the transcription factor linked to a membrane anchoring domain via a protease cleavage site; and

(iii)   detecting in each cell whether the altered protease has cleaved the protease cleavage site to release the transcription factor by measuring expression of the target gene.

11.   A method for discovering protease genes, which method comprises:

(i)   creating a gene construct which codes for a transcription factor precursor, which comprises a transcription factor linked to a membrane anchoring domain via a protease cleavage site;

- 49 -

(ii)   introducing the gene construct into a range of different cells which contain a target gene under the control of the transcription factor;

(iii)   detecting in which cells the protease is expressed by determining whether any protease has cleaved the protease cleavage site to release the transcription factor by measuring expression of the target gene; and

(iv)   recovering and sequencing the protease gene from cells which express the target gene.

12.   A method as claimed in any claim from 8 to 10 wherein the membrane anchoring domain and protease cleavage site are not derived from the same protein.
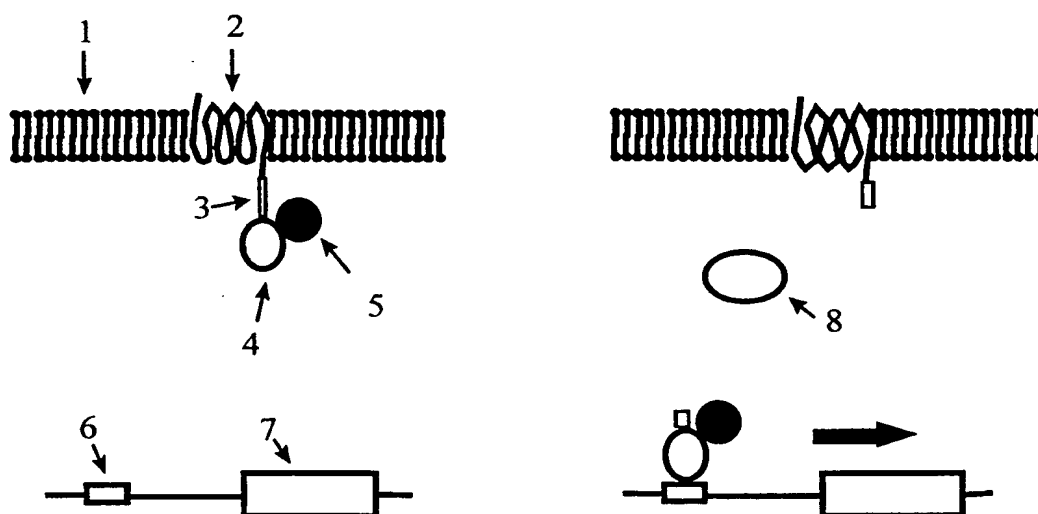
Fig.1

Fig. 2

LDD208   **TEV protease**

LDD855   **TEV protease [ΔC-16]**

LDD859   **TEV protease [ΔI-10]**

Fig. 3